Designing various component analysis at will

Akisato Kimura, Hitoshi Sakano, Hirokazu Kameoka
NTT Communication Science Laboratories
akisato@ieee.org

Masashi Sugiyama
Tokyo Institute of Technology
sugi@cs.titech.ac.jp

Abstract

This paper provides a generic framework of component analysis (CA) methods introducing a new expression for scatter matrices and Gram matrices, called Generalized Pairwise Expression (GPE). This expression is quite compact but highly powerful: The framework includes not only (1) the standard CA methods but also (2) several regularization techniques, (3) weighted extensions, (4) some clustering methods, and (5) their semi-supervised extensions. This paper also presents quite a simple methodology for designing a desired CA method from the proposed framework: Adopting the known GPEs as templates, and generating a new method by combining these templates appropriately.

1 Introduction

Component analysis (CA) is traditional, quite simple but might be one of the powerful tools to obtain a hidden structure embedded in the data. Recent reports showed its effectiveness for several tasks in computer vision and pattern recognition. Principal component analysis (PCA), Fisher discriminant analysis (FDA), multiple linear regression (MLR), and canonical correlation analysis (CCA) are well known as standard CA methods [1]. They can be formulated as a generalized eigenvalue problem of a scatter matrix or an augmented matrix composed of several scatter matrices [2, 3]. Kernel CA methods as kernelized extensions of those standard methods have been also developed to deal with non-vector samples and non-linear analysis, which can be formulated as a generalized eigenvalue problem of Gram matrices, instead of scatter matrices [2, 3]. Kernel CA often needs some regularization techniques such as ℓ_2-norm regularization to inhibit overfitting and Laplacian regularization to fit underlying data manifolds smoothly. In addition, improvements of robustness against outliers and separately distributed samples and their extensions to semi-supervised analysis have been considered.

Although a lot of CA methods and several trials to unify these methods have been presented so far [2, 3, 4, 5], freely designing a tailor-made method of CA for a specific purpose or domain still remains an open problem. Until now, researchers have had to choose one of the existing methods that seems best to address the problem of interest, or had to laboriously develop a new analysis method tailored specifically for that purpose.

In view of the above discussions, this paper provides a new expression of scatter matrices and Gram matrices, which we call *generalized pairwise expression (GPE)* to make it easy to design a new CA method with desired properties. The methodology is quite simple: **adopting the above mentioned special cases as templates, and generating a new method by combining these templates appropriately**. This characteristics has not been discussed yet in any previous researches to our best knowledge. It is also possible to individually select and arrange samples for calculating the scatter matrices of the methods to be combined, which enables us to extend CA methods to semi-supervised ones and multi-modal ones.

References

[1] T. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, 2003.

[2] M. Borga+. A unified approach to PCA, PLS, MLR and CCA. Technical Report LiTH-ISY-R-1992, Linkoping University, 1997.

[3] T. De Bie+. Eigenproblems in pattern recognition. In *Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics*, Springer, 2005.

[4] F. De la Torre. A unification of component analysis methods. In *Handbook of Pattern Recognition and Computer Vision*, World Scientific Pub, 2010.

[5] F. De la Torre. A least-squares framework for component analysis. *IEEE Trans. PAMI*, 2012.